CLAIMS

What is claimed is:

1. A computer-implemented method for identifying optimal allocations of computing resources in a data processing arrangement having a plurality of computing machines that host a plurality of application processes, comprising:

establishing a plurality of server models, each server model including one or more server nodes, wherein each server node has an associated set of capacity attributes;

designating a layered relationship between the server models, wherein for a first server-model layer immediately above a second server-model layer, the second server-model layer includes respective models that represent the nodes in the first server-model layer;

establishing a plurality of service models, each service model including one or more service nodes, wherein each service node has an associated set of demand attributes;

designating a layered relationship between the service models, wherein for a first service-model layer immediately above a second service-model layer, the second service-model layer includes respective models that represent the nodes in the first server-model layer; and

generating an optimized mapping of service nodes in a user-selected service model to server nodes in a user-selected server model as a function of the demand and capacity attributes.

2. The method of claim 1, further comprising:

monitoring, while the applications processes are executing, levels of demand for computing resources that correspond to selected ones of the demand attributes;

storing the levels of demand; and

generating an alternate optimized mapping of service nodes in a user-selected service model to server nodes in a user-selected server model using the stored levels of demand and the capacity attributes.

3. The method of claim 2, further comprising:

establishing one or more service-node relationships between selected pairs of the service nodes, wherein each service-node relationship has an associated transport demand attribute specifying a quantity of communication resources required for communication between the associated pair of service nodes;

establishing one or more server-node relationships between selected pairs of the server nodes, wherein each server-node relationship has an associated transport capacity attribute specifying a quantity of communication resources available for communication between the associated pair of server nodes; and

generating the optimized mapping as a function of the service-node relationships and server-node relationships.

4. The method of claim 3, wherein each service node has an associated set of capacity attributes and further comprising generating an optimized mapping of service nodes in a first user-selected service model to service nodes in a second user-selected service model as a function of the demand attributes of the first service model and capacity attributes of the second service model.

5. The method of claim 4, wherein each server node has an associated set of demand attributes and further comprising generating an optimized mapping of server nodes in a first user-selected server model to server nodes in a second user-selected server model as a function of the demand attributes of the first server model and capacity attributes of the second server model.

6. The method of claim 5, further comprising:

representing the service models and server models in XML; and

generating a allocation matrix in XML that represents the optimized mapping.

7. The method of claim 1, further comprising:

establishing one or more service-node relationships between selected pairs of the service nodes, wherein each service-node relationship has an associated transport demand

attribute specifying a quantity of communication resources required for communication between the associated pair of service nodes;

establishing one or more server-node relationships between selected pairs of the server nodes, wherein each server-node relationship has an associated transport capacity attribute specifying a quantity of communication resources available for communication between the associated pair of server nodes; and

generating the optimized mapping as a function of the service-node relationships and server-node relationships.

8. The method of claim 7, wherein each service node has an associated set of capacity attributes and further comprising generating an optimized mapping of service nodes in a first user-selected service model to service nodes in a second user-selected service model as a function of the demand attributes of the first service model and capacity attributes of the second service model.

9. The method of claim 1, wherein each service node has an associated set of capacity attributes and further comprising generating an optimized mapping of service nodes in a first user-selected service model to service nodes in a second user-selected service model as a function of the demand attributes of the first service model and capacity attributes of the second service model.

10. The method of claim 9, wherein each server node has an associated set of demand attributes and further comprising generating an optimized mapping of server nodes in a first user-selected server model to server nodes in a second user-selected server model as a function of the demand attributes of the first server model and capacity attributes of the second server model.

11. The method of claim 1, wherein each server node has an associated set of demand attributes and further comprising generating an optimized mapping of server nodes in a first user-selected server model to server nodes in a second user-selected server model as a function of the demand attributes of the first server model and capacity attributes of the second server model.

12. An apparatus for identifying optimal allocations of computing resources in a data processing arrangement having a plurality of computing machines that host a plurality of application processes, comprising:

means for establishing a plurality of server models, each server model including one or more server nodes, wherein each server node has an associated capacity attribute;

means for designating a layered relationship between the server models, wherein for a first server-model layer immediately above a second server-model layer, the second server-model layer includes respective models that represent the nodes in the first server-model layer;

means for establishing a plurality of service models, each service model including one or more service nodes, wherein each service node has an associated demand attribute;

means for designating a layered relationship between the service models, wherein for a first service-model layer immediately above a second service-model layer, the second service-model layer includes respective models that represent the nodes in the first server-model layer; and

means for generating an optimized mapping of service nodes in a user-selected service model to server nodes in a user-selected server model as a function of the demand and capacity attributes.

13. A system for identifying optimal allocations of computing resources in a data processing arrangement having a plurality of computing machines that host a plurality of application processes, comprising:

a model repository including a plurality of server models and a plurality of service models, each server model including one or more server nodes and each server node having an associated set of capacity attributes, each service model including one or more service nodes and each service node having an associated set of demand attributes, wherein the server models are defined in a layered relationship and for a first server-model layer immediately above a second server-model layer, the second server-model layer includes respective models that represent the nodes in the first server-model layer, and the service models are defined in a layered relationship and for a first service-model layer

immediately above a second service-model layer, the second service-model layer includes respective models that represent the nodes in the first service-model layer; and

an optimization engine coupled to the model repository, the optimization engine configured to generate an optimized mapping of service nodes in a user-selected service model to server nodes in a user-selected server model as a function of the associated demand and capacity attributes.

14. The system of claim 13, further comprising:

means for monitoring, while the applications processes are executing, levels of demand for computing resources that correspond to selected ones of the demand attributes;

means for storing the levels of demand; and

wherein the optimization engine is further configured to generate an alternate optimized mapping of service nodes in a user-selected service model to server nodes in a user-selected server model using the stored levels of demand and the capacity attributes.

15. The system of claim 13, further comprising:

wherein the model repository further includes one or more service-node relationships between selected pairs of the service nodes, each service-node relationship having an associated transport demand attribute that specifies a quantity of communication resources required for communication between the associated pair of service nodes;

wherein the model repository further includes one or more server-node relationships between selected pairs of the server nodes, each server-node relationship having an associated transport capacity attribute that specifies a quantity of communication resources available for communication between the associated pair of server nodes; and

the optimization engine is further configured to generate the optimized mapping as a function of the service-node relationships and server-node relationships.

16. The system of claim 13, wherein each service node has an associated set of capacity attributes and the optimization engine is further configured to generate an optimized mapping of service nodes in a first user-selected service model to service nodes

in a second user-selected service model as a function of the demand attributes of the first service model and capacity attributes of the second service model.

17. The system of claim 13, wherein each server node has an associated set of demand attributes and the optimization engine is further configured to generate an optimized mapping of server nodes in a first user-selected server model to server nodes in a second user-selected server model as a function of the demand attributes of the first server model and capacity attributes of the second server model.